



## SEQUENCE LISTING



<110> Tomikawa, Mayumi  
Aikawa, Seiichi  
Matsuzawa, Fumiko

<120> Method and Apparatus for Extracting and Evaluating Mutually Similar Portions in One-Dimensional Sequences in Molecules and/or Three-Dimensional Structures of Molecules

<130> 522.1921D3

<140> 09/909,809  
<141> 2001-07-23

<150> US 08/014,867  
<151> 1993-02-08

<160> 20

<170> PatentIn version 3.1 (created in U.S. Serial No. 09/910,071)

<210> 1  
<211> 37  
<212> PRT  
<213> human

<400> 1

Gly Asp Val Glu Lys Gly Lys Lys Ile Phe Ile Met Lys Cys Ser Gln  
1 5 10 15

His Thr Val Glu Gly Gly Lys His Lys Thr Gly Pro Asn Leu His Gly  
20 25 30

Leu Phe Gly Arg Lys  
35

<210> 2  
<211> 39  
<212> PRT  
<213> bacterium

<400> 2

Glu Gly Asp Asp Ala Ala Ala Gly Glu Lys Val Ser Lys Lys Cys Leu  
1 5 10 15

Ala Cys His Thr Phe Asp Gln Gly Gly Ala Asn Lys Val Gly Pro Asn  
20 25 30

Pro Asn Leu Phe Gly Val Phe

<210> 3  
<211> 51  
<212> PRT  
<213> rat

<400> 3

Met Ser Leu Ala Ile Leu Arg Val Ile Arg Leu Val Arg Val Phe Arg  
1 5 10 15

Ile Phe Lys Leu Ser Arg His Ser Lys Gly Leu Gln Ile Leu Gly Arg  
20 25 30

Thr Leu Lys Ala Ser Met Arg Glu Leu Gly Leu Leu Ile Phe Phe Ile  
35 40 45

Gly Val Val  
50

<210> 4  
<211> 142  
<212> PRT  
<213> unknown

<220>  
<223> Figure 23A

<400> 4

Thr Glu Glu Gln Ile Ala Glu Phe Lys Glu Ala Phe Ser Leu Phe Asp  
1 5 10 15

Lys Asp Gly Asp Gly Thr Ile Thr Thr Lys Glu Leu Gly Thr Val Met  
20 25 30

Arg Ser Leu Gly Gln Asn Pro Thr Glu Ala Glu Leu Gln Asp Met Ile  
35 40 45

Asn Glu Val Asp Ala Asp Gly Asn Gly Thr Ile Asp Phe Pro Glu Phe  
50 55 60

Leu Thr Met Met Ala Arg Lys Met Lys Asp Thr Ser Glu Glu Glu Ile  
65 70 75 80

Arg Glu Ala Phe Arg Val Phe Asp Lys Asp Gly Asn Gly Tyr Ile Ser

85

90

95

Ala Ala Glu Leu Arg His Val Met Thr Asn Leu Gly Glu Lys Leu Thr  
100 105 110

Asp Glu Glu Val Asp Glu Met Ile Arg Glu Ala Asn Ile Asp Gly Asp  
115 120 125

Gly Gln Val Asn Tyr Glu Glu Phe Val Gln Met Met Thr Ala  
130 135 140

<210> 5  
<211> 159  
<212> PRT  
<213> unknown

<220>  
<223> Figure 23B

<400> 5

Ala Met Asp Gln Gln Ala Glu Ala Arg Ala Phe Leu Ser Glu Glu Met  
1 5 10 15

Ile Ala Glu Phe Lys Ala Ala Phe Asp Met Phe Asp Ala Asp Gly Gly  
20 25 30

Gly Asp Ile Ser Thr Lys Glu Leu Gly Thr Val Met Arg Met Leu Gly  
35 40 45

Gln Asn Pro Thr Lys Glu Glu Leu Asp Ala Ile Ile Glu Glu Val Asp  
50 55 60

Glu Asp Gly Ser Gly Thr Ile Asp Phe Glu Glu Phe Leu Val Met Val  
65 70 75 80

Arg Gln Met Lys Glu Asp Ala Lys Gly Lys Ser Glu Glu Glu Leu Ala  
85 90 95

Asp Cys Phe Arg Ile Phe Asp Lys Asn Ala Asp Gly Phe Ile Asp Ile  
100 105 110

Glu Glu Leu Gly Glu Ile Leu Arg Ala Thr Gly Glu His Val Thr Glu  
115 120 125

Glu Asp Ile Glu Asp Leu Met Lys Asp Ser Asp Lys Asn Asn Asp Gly  
130 135 140

Arg Ile Asp Phe Asp Glu Phe Leu Lys Met Met Glu Gly Val Gln  
145 150 155

<210> 6  
<211> 28  
<212> PRT  
<213> unknown

<220>  
<223> calmodulin probe site 81-108, Figure 25, target

<400> 6

Leu Ala Asp Cys Phe Arg Ile Phe Asp Lys Asn Ala Asp Gly Phe Ile  
1 5 10 15

Asp Ile Glu Glu Leu Gly Glu Ile Leu Arg Ala Thr  
20 25

<210> 7  
<211> 28  
<212> PRT  
<213> unknown

<220>  
<223> calmodulin probe site 81-108, probe, Figure 25

<400> 7

Ile Arg Glu Ala Phe Arg Val Phe Asp Lys Asp Gly Asn Gly Tyr Ile  
1 5 10 15

Ser Ala Ala Glu Leu Arg His Val Met Thr Asn Leu  
20 25

<210> 8  
<211> 27  
<212> PRT  
<213> unknown

<220>  
<223> Calmodulin probe site 117-143, Figure 26, target

<400> 8

Ile Glu Asp Leu Met Lys Asp Ser Asp Lys Asn Asn Asp Gly Arg Ile  
1 5 10 15

Asp Phe Asp Glu Phe Leu Lys Met Met Glu Gly  
20 25

<210> 9  
<211> 27  
<212> PRT  
<213> unknown  
  
<400> 9

Val Asp Glu Met Ile Arg Glu Ala Asn Ile Asp Gly Asp Gly Gln Val  
1 5 10 15

Asn Tyr Glu Glu Phe Val Gln Met Met Thr Ala  
20 25

<210> 10  
<211> 8  
<212> PRT  
<213> unknown

<220>  
<223> Coordinates of C-alpha corresponding to the aminon acid residue N  
os. 7 to 14 in elongation factor of protein which is a binding si  
te for phosphoric acid of GTP

<400> 10

Gly His Val Asp His Gly Lys Thr  
1 5

<210> 11  
<211> 8  
<212> PRT  
<213> unknown

<220>  
<223> adenylate kinase - target

<400> 11

Gly Ala Pro Gly Ser Gly Lys Gly  
1 5

<210> 12  
<211> 8  
<212> PRT  
<213> unknown

<220>  
<223> ras protein - target

<400> 12

Gly Ala Gly Gly Val Gly Lys Ser  
1 5

<210> 13

<211> 8  
<212> PRT  
<213> unknown

<220>

<223> adenylate kinase (3ADK)

<400> 13

Gly Gly Pro Gly Ser Gly Lys Gly  
1 5

<210> 14

<211> 223  
<212> PRT  
<213> unknown

<220>

<223> amino acid sequence of trypsin

<400> 14

Ile Val Gly Gly Tyr Thr Cys Cys Ala Asn Thr Val Pro Tyr Gln Val  
1 5 10 15

Ser Leu Asn Ser Gly Tyr His Phe Cys Gly Gly Ser Leu Ile Asn Ser  
20 25 30

Gln Trp Val Val Ser Ala Ala His Cys Tyr Lys Ser Gly Ile Gln Val  
35 40 45

Arg Leu Gly Glu Asp Asn Ile Asn Val Val Glu Gly Asn Glu Gln Phe  
50 55 60

Ile Ser Ala Ser Lys Ser Ile Val His Pro Ser Tyr Asn Ser Asn Thr  
65 70 75 80

Leu Asn Asn Asp Ile Met Leu Ile Lys Leu Lys Ser Ala Ala Ser Leu  
85 90 95

Asn Ser Arg Val Ala Ser Ile Ser Leu Pro Thr Ser Cys Ala Ser Ala  
100 105 110

Gly Thr Gln Cys Leu Ile Ser Gly Trp Gly Asn Thr Lys Ser Ser Gly  
115 120 125

Thr Ser Tyr Pro Asp Val Leu Lys Cys Leu Lys Ala Pro Ile Leu Ser  
130 135 140

Asp Ser Ser Cys Lys Ser Ala Tyr Pro Gly Gln Ile Thr Ser Asn Met  
145 150 155 160

Phe Cys Ala Gly Tyr Leu Glu Gly Lys Asp Ser Cys Gln Gly Asp  
165 170 175

Ser Gly Gly Pro Val Val Cys Ser Gly Lys Leu Gln Gly Ile Val Ser  
180 185 190

Trp Gly Ser Gly Cys Ala Gln Lys Asn Lys Pro Gly Val Tyr Thr Lys  
195 200 205

Val Cys Asn Tyr Val Ser Trp Ile Lys Gln Thr Ile Ala Ser Asn  
210 215 220

<210> 15

<211> 239

<212> PRT

<213> unknown

<220>

<223> amino acid sequence of Elastase

<400> 15

Val Val Gly Gly Thr Glu Ala Gln Arg Asn Ser Trp Pro Ser Gln Ile  
1 5 10 15

Ser Leu Gln Tyr Arg Ser Gly Ser Ser Trp Ala His Thr Cys Gly Gly  
20 25 30

Thr Leu Ile Arg Gln Asn Trp Val Met Thr Ala Ala His Cys Val Asp  
35 40 45

Arg Glu Leu Thr Phe Arg Val Val Gly Glu His Asn Leu Asn Gln  
50 55 60

Asn Asn Gly Thr Glu Gln Tyr Val Gly Val Gln Lys Ile Val Val Pro

65

70

75

80

Tyr Trp Asn Thr Asp Asp Val Ala Ala Gly Tyr Asp Ile Ala Leu Leu  
85 90 95

Arg Leu Ala Gln Ser Val Thr Leu Asn Ser Tyr Val Gln Leu Gly Val  
100 105 110

Leu Pro Arg Ala Gly Thr Ile Leu Ala Asn Ser Pro Cys Tyr Ile Thr  
115 120 125

Thr Gly Trp Gly Leu Thr Arg Thr Asn Gly Gln Leu Ala Gln Thr Leu  
130 135 140

Gln Gln Ala Tyr Leu Pro Thr Val Asp Tyr Ala Ile Cys Ser Ser Ser  
145 150 155 160

Ser Tyr Trp Gly Ser Thr Val Lys Asn Ser Met Val Cys Ala Gly Gly  
165 170 175

Asp Gly Val Arg Ser Gly Cys Gln Gly Asp Ser Gly Gly Pro Leu His  
180 185 190

Cys Leu Val Asn Gly Gln Tyr Ala Val His Gly Val Thr Ser Phe Val  
195 200 205

Ser Arg Leu Gly Cys Asn Val Thr Arg Lys Pro Thr Val Phe Thr Arg  
210 215 220

Val Ser Ala Tyr Ile Ser Trp Ile Asn Asn Val Ile Ala Ser Asn  
225 230 235

<210> 16

<211> 6

<212> PRT

<213> unknown

<220>

<223> histidine active sites (41-46) of elastase

<400> 16

Met Thr Ala Ala His Cys  
1 5

<210> 17  
<211> 6  
<212> PRT  
<213> unknown  
  
<220>  
<223> histidine active sites (36-41) of trypsin  
  
<400> 17

Val Ser Ala Ala His Cys  
1 5

<210> 18  
<211> 5  
<212> PRT  
<213> unknown  
  
<220>  
<223> serine active sites of elastase  
  
<400> 18

Gly Asp Ser Gly Gly  
1 5

<210> 19  
<211> 5  
<212> PRT  
<213> unknown  
  
<220>  
<223> serine active sites (175-179) of trypsin  
  
<400> 19

Gly Asp Ser Gly Gly  
1 5

<210> 20  
<211> 120  
<212> PRT  
<213> Artificial Sequence  
  
<220>  
<223> Figure 46 - examples of outputs protein A and protein B  
  
<400> 20

Ala Thr Tyr Arg Leu Lys Asn Gly Asp Asp Pro Leu Ser Asp His Lys  
1 5 10 15

Leu Ser Asp Phe Gly Phe Ser Gly Lys Ile Leu Thr Arg Tyr Asn Gln

20

25

30

Ala Val Gly Asp Asp Pro Leu Asp Gly Ala Arg Ser Asp Phe Gly Phe  
35 40 45

Ser Gly Lys Ile Pro Leu Thr Arg Tyr Asn Val Gln Asn Asp Glu Ser  
50 55 60

Gly Phe Arg Thr Lys His Gly Leu Asp Ala Ser Asn Ala Ala Thr Tyr  
65 70 75 80

Arg Leu Lys Asn Asp Asp Pro Leu Asp Gly Ala Arg Glu Lys Leu Ser  
85 90 95

Asp Phe Gly Ala His Phe Ser Gly Lys Ile Phe Thr Arg Tyr Asn Gln  
100 105 110

Ala Val Gln Asn Asp Glu Ser Lys  
115 120